

**CEDE****DOCUMENTO CEDE 2005-31**
ISSN 1657-7191 (Edición Electrónica)
MAYO DE 2005

EL CENSO NACIONAL DE POBLACIÓN: UNA COMPARACIÓN DE METODOLOGÍAS MEDIANTE SIMULACIONES DE MONTE CARLO[†]

CHRISTIAN JARAMILLO^{*}
ANA MARÍA IBÁÑEZ^{}**

Resumen

Bajo una fuerte polémica, Colombia se prepara para aplicar un nuevo censo de población. La adopción de una enumeración continua constituye una de las principales modificaciones y es el centro de la polémica suscitada. Numerosas dudas han surgido en torno al nuevo diseño y un interrogante particular es la efectividad de un censo continuo dado la alta movilidad de la población colombiana debido al fenómeno del desplazamiento forzado. El objetivo de este artículo es evaluar la exactitud de la información obtenida y su degradación, es decir, la pérdida de exactitud en el conteo de la población colombiana a través del tiempo para los dos tipos de censos: censo tradicional y censo continuo. La comparación, basada en simulaciones de Monte Carlo, indica que la diferencia en la exactitud y degradación de la información entre los dos tipos de censo, aunque estadísticamente significativa, es mínima.

Palabras clave: Censo Nacional de Población, simulaciones de Monte Carlo, desplazamiento forzado

Clasificación JEL: C15, J10, J60

[†] Los autores agradecen a Bibiana Taboada la ayuda prestada en la programación y el análisis de las simulaciones.

^{*} Profesor Asistente. Facultad de Economía. Universidad de los Andes. chjarami@uniandes.edu.co

^{**} Profesora Asistente. Facultad de Economía. Universidad de los Andes. aibanez@uniandes.edu.co

THE NACIONAL POPULATION CENSUS: A COMPARISON OF ENUMERATION METHODS USING MONTE CARLO SIMULATIONS

Abstract

After a strong debate, Colombia is preparing to administer a new population census. The adoption of a continuous enumeration is one of the main features of the new census and the crux of the debate. Given the high mobility of the Colombian population as a result of forced displacement, questions about the effectiveness of a continuous census have arisen. The purpose of this paper is to evaluate the accuracy of the information elicited by the census and its deterioration rate over time, that is, the loss in accuracy of the enumeration of population over time for both types of census: the traditional census and the continuous census. The results, based on Monte Carlo simulations, indicate that the difference in accuracy and the deterioration of the quality of information between both censuses, although statistically significant, is small.

Key words: National Population Census, Monte Carlo simulations, forced displacement

JEL Classification: C15, J10, J60

I. Introducción

Un censo de población es el proceso de recolectar, compilar, evaluar y analizar información demográfica, económica y social de la totalidad de habitantes de un país en un momento determinado del tiempo. Los censos son importantes instrumentos para la toma de decisiones, la planeación y la administración pública. Al proporcionar un diagnóstico del estado de la población los censos permiten asignar recursos de inversión de manera eficiente; proveen información para evaluar y ejecutar programas sociales; y son la base para elaborar encuestas de muestreo. Los censos, además, son herramientas esenciales de investigación académica y suministran información útil para estudios de mercado (ONU, 1997).

Colombia se prepara para iniciar un nuevo censo de población bajo una fuerte polémica. El Gobierno Nacional decidió modificar el proceso de enumeración tradicionalmente utilizado en el Censo y extendió el periodo de enumeración al adoptar la metodología de un censo continuo en la cual se inmovilizan las regiones en distintos momentos del tiempo y no de manera simultánea.

Diversos argumentos se han esgrimido en contra del nuevo diseño del censo. El riesgo de un conteo erróneo de la población debido a la alta movilidad de la población colombiana, consecuencia del desplazamiento forzado, es mencionado con frecuencia. En los últimos diez años, el desplazamiento forzado ha significado la movilización de 1'542.915 personas¹, lo cual equivale a un poco más de tres por ciento de la población colombiana y 12 por ciento de la población rural. Ello ha conducido a una recomposición de la población considerable en municipios pequeños y en ciudades intermedias. Por ejemplo, la intensidad del desplazamiento, indicador que refleja el impacto de la expulsión de población en los municipios, en San Carlos, Antioquia y El Carmen de Bolívar equivale a 72.739 y 69.766 por cada 100.000 habitantes, respectivamente. Asimismo, los flujos de población en las ciudades intermedias alcanzan niveles preocupantes con

¹ Red de Solidaridad Social, www.red.gov.co.

presiones del desplazamiento² en Quibdó, Sincelejo y Florencia de 26.4 por ciento, 24.6 por ciento y 20.3 por ciento respectivamente.

El objetivo de este artículo es comparar el desempeño en el conteo de población de los dos tipos de censo – censo continuo y censo tradicional - dado que existe en la actualidad una alta movilidad de la población colombiana. Con el fin de realizar esta comparación, se estima el error de conteo de ambos censos y se incorpora en la estimación una matriz de movilidad poblacional basada en los datos de población desplazada. Asimismo, se examina el deterioro de la información de ambos censos a lo largo del tiempo. Cabe anotar que este artículo, al estimar únicamente el error en el conteo de población, no pretende en ningún momento embarcarse en una comparación exhaustiva de ambos tipos de censo.

Los resultados muestran que el censo continuo, comparado con el tradicional, no implica pérdidas significativas en la calidad de la información demográfica. El error regional porcentual promedio y su variabilidad son de magnitudes similares en ambos tipos de censos. En particular, la inexactitud de un censo tradicional tras un año de realizado es mayor que la diferencia en exactitud entre dos censos continuos realizados con un año de diferencia. Este resultado no varía en presencia de crecimiento poblacional ni de flujos imprevistos de población.

El artículo está compuesto por tres secciones. La primera sección describe el fenómeno del desplazamiento forzoso en Colombia y discute las posibles implicaciones de la migración forzada sobre el desempeño de los dos tipos de censos. La segunda sección explica metodología usada para simular el desempeño de los dos tipos de censo – censo tradicional y censo continuo, describe los datos y discute los resultados. Las conclusiones e implicaciones de política se discuten en la tercera sección.

² La presión del desplazamiento se calcula como el porcentaje de desplazados que arriban al municipio receptor respecto a la población nativa.

II. El desplazamiento forzoso en Colombia: ¿Cuál es el efecto del desplazamiento forzoso sobre el desempeño estadístico de los censos?

Durante la última década, los actores armados han arreciado sus ataques a la población civil con masacres, asesinatos selectivos, reclutamiento forzosos y confrontaciones armadas. Ello ha ocasionado la expulsión masiva de población civil de sus municipios de origen y, en el periodo comprendido entre 1995 a 2005, un poco más de 1'542.915 de colombianos fueron víctimas del desplazamiento forzoso. El grueso del desplazamiento, cerca del 64 por ciento, se produjo durante los años 2001 a 2003. Además, el desplazamiento se expande a lo largo y ancho del territorio nacional. Según datos de la RSS, la gran mayoría de los municipios del país son receptores o expulsores de población: 997 son receptores y 1.071 son expulsores de población desplazada. Los principales departamentos receptores de población desplazada son Antioquia, Bolívar y Sucre y los principales departamentos expulsores son Antioquia, Bolívar y Magdalena.

El desplazamiento se produce como consecuencia de un ataque de los grupos armados, ya sea directamente sobre la familia o sobre personas cercanas. Hay distintos tipos de desplazamiento. Por un lado, algunos hogares se trasladan al interior de sus municipios, por lo general, del área rural a la cabecera municipal. Dichos desplazamientos intermunicipales suelen ser temporales ya que los hogares están esperando que la situación de orden público se estabilice para retornar a sus parcelas. Por otro lado, algunos hogares migran de un municipio a otro, ya sea en el mismo departamento o en otro departamento. Los desplazamientos intradepartamentales suelen ser permanentes o de un periodo más prolongado en contraste con los intramunicipales. Cerca de 60.1 por ciento del desplazamiento es intradepartamental y 25.6 por ciento es intramunicipal (Ibáñez y Querubín, 2004). Asimismo, el desplazamiento puede ser individual, es decir el hogar migra sólo, o masivo³, es decir el hogar migra con un grupo de

³ El Decreto 2569 de 2000 define el desplazamiento masivo como la migración forzada que afecta, en una misma circunstancia de tiempo, modo y lugar, a 50 o más personas o a 10 o más hogares.

hogares. En Colombia tres cuartas partes de los desplazamientos son individuales⁴. Por último, cerca de 95 por ciento de las familias migran directamente del municipio de origen al municipio receptor (Ibáñez y Querubín, 2004).

El desplazamiento forzoso dificulta el proceso de enumeración de la población y puede causar errores en el conteo. El diseño inicial del censo, al inmovilizar la población en un mismo día y enumerar el número de habitantes en ese momento del tiempo, evita que se presenten duplicaciones u omisiones en el proceso de conteo de la población. El censo continuo, al inmovilizar cada región en lapsos distintos del tiempo, incrementa la probabilidad de duplicaciones u omisiones. Ello redundaría en un conteo erróneo de la población. Es importante, por ende, establecer, dado los flujos de desplazamiento forzado, el error de conteo de población para cada uno de los censos - censo tradicional y censo continuo- y comparar así su desempeño. Asimismo, se debe examinar como varía el desempeño regional de cada censo, información valiosa para diseñar el censo de población de tal manera que se minimice el error de conteo. La sección siguiente simula el desempeño de los dos censos y compara los errores de conteo.

⁴ Red de Solidaridad Social, www.red.gov.co.

III. Simulaciones.

El objetivo de esta sección es evaluar la efectividad de los dos tipos de censo para enumerar la población colombiana. Para alcanzar tal objetivo, se llevan a cabo simulaciones de Monte Carlo que representan la evolución de la población colombiana en un periodo de diez años. Para efectos de referencia, la población simulada se considera la "real". En cada simulación se cuenta la población con los dos tipos de censo y se describen las discrepancias entre los conteos y la población real.

Con el fin de establecer el efecto del crecimiento poblacional y del desplazamiento forzoso sobre el desempeño del censo, se llevan a cabo dos grupos de simulaciones. El primer grupo, cuyo principal propósito es observar los efectos del crecimiento demográfico y su interacción con los flujos poblacionales, independientemente de la configuración inicial de la población. El segundo grupo realiza un ejercicio similar para el caso particular de la población colombiana, es decir adapta las simulaciones al tamaño de la población colombiana, su distribución regional y los flujos poblacionales entre regiones.

3.1. Descripción de la metodología

3.1.1 Simulación de Monte Carlo

El proceso que se lleva a cabo parte de un vector de población inicial (x^0) que contiene información para cada una de las 48 regiones en que se subdivide el país. Este vector cambia, para periodos siguientes, de acuerdo con el comportamiento de dos variables, a saber: el vector de crecimiento demográfico (g), entendido como natalidad neta de mortalidad; y la matriz de desplazamientos entre regiones (D^t). Si bien estos dos factores influyen en la evolución del vector poblacional, es en el movimiento de individuos entre regiones donde se centra el interés del estudio. Por esta razón, mientras que el parámetro de crecimiento poblacional se mantiene constante para toda una simulación, los desplazamientos se simulan estocásticamente mes a mes.

Para las simulaciones de la matriz de desplazamientos mensuales (D^t) se utiliza como punto de partida la matriz de transferencias (T). Esta última se basa en un registro histórico de los movimientos de individuos entre regiones en Colombia: cada t_{ij} corresponde al número de emigrantes mensuales promedio de la región i a la región j en los últimos cinco años. En la simulación, cada d_{ij}^t es el número de individuos que se desplazan de la región i a la región j en el mes t , y su valor es una realización particular de una variable estocástica con distribución de Poisson y valor esperado t_{ij} .

La ecuación con base en la cual se actualiza el vector de población es la siguiente:

$$x^{t+1} = (1 + g) * \left[x^{t+1} + (D^t)^T * \bar{1} - D^t * \bar{1} \right]$$

donde $\bar{1}$ es un vector de unos y el superíndice T indica una matriz traspuesta.

Partiendo de que en Colombia se realizan censos cada 10 años, y de que un censo continuo dura un año, cada simulación se hace para 132 periodos. Esto corresponde al lapso de tiempo entre el comienzo del primer censo continuo y el final del siguiente, es decir, a once años.

Para estimar la medición de los dos tipos de censo, se selecciona la información que prescribe la metodología en cada uno de los casos y se guarda en vectores determinados (c_k).

Para el primer censo tradicional se guarda, en los vectores c_1 y c_3 , la información de cada una de las regiones en el periodo 1 (primer mes del año 1) y en el periodo 12 (último mes del año 1), respectivamente. Similarmente, para el segundo censo de este tipo, se repite el proceso en los periodos 120 y 132 y se almacenan los datos en los vectores c_4 y c_6 .

El proceso es menos simple para el caso del censo continuo ya que la metodología que éste emplea exige que las mediciones se realicen durante todos los meses del año en el que se efectúa el censo. Es necesario agregar

información al vector en el que se guardan los datos en cada uno de los meses del año. La manera en la que se van llenando los vectores c_2 y c_5 , para los años 1 y 10, respectivamente, es la siguiente: en el primer mes de cada año censado se guarda la información correspondiente a las primeras cuatro regiones, en el segundo a las siguientes cuatro, y así sucesivamente hasta llegar último mes del año censado en el que se guarda la información de las últimas cuatro regiones. Así, al final de cada uno de los años censados, se completa un vector con información para cada una de las regiones del país.

Teniendo en cuenta que la vigencia aproximada de los censos es de diez años, también es necesario evaluar su exactitud durante este periodo de tiempo. Es decir, resulta oportuno saber en qué medida una medición que se lleva a cabo en un año sigue siendo válida para los siguientes diez, hasta que se realice una nueva medición. Con este fin se guarda la información del vector de población real al final de cada año y se compara con la información registrada en los censos.

Los procesos descritos corresponden a mediciones de una iteración de la simulación. Ésta se repite 1000 veces y en cada repetición se guarda la información mencionada. De esta forma es posible obtener estadísticas sobre el desempeño de los censos bajo patrones de desplazamiento factibles. Con el fin de analizar las características y la evolución del error regional a través del tiempo, se definen medidas de error porcentual promedio y varianza del error.

3.1.2 Medidas de error

La distinción entre los datos globales y los regionales es crucial porque el error global posiblemente subestima el desfase a nivel de las regiones.⁵ En la medida en que el censo tiene implicaciones para políticas redistributivas, el dato regional adquiere mayor relevancia.

⁵ Por ejemplo, en el caso en que el crecimiento poblacional es cero, el error agregado es necesariamente cero en el censo tradicional, pero la distribución de la población entre regiones puede variar con respecto a lo que el censo dice.

Como medidas del error global en el conteo de la población se calculan el error promedio en el total de la población para cada periodo t , en términos absolutos (EGt) y porcentuales ($EGPt$), así como su desviación estándar ($DEGPt$).

Para el análisis del error de la medición a nivel de las regiones se utilizan dos medidas. La primera es el error porcentual promedio en el periodo t , definido para el censo c como:

$$EPP_t = \frac{1}{N} \sum_{i \in N} \frac{abs(x_{it} - c_i)}{x_{it}}$$

La segunda medida es la desviación estándar del error porcentual en cada periodo t :

$$DEP_t = \frac{1}{N(N-1)} \sum_{i \in N} \left(\frac{abs(x_{it} - c_i)}{x_{it}} - EPP_t \right)^2$$

Mientras que el $EPPt$ es una medida de la magnitud del desfase entre el dato del censo y la población real, la $DEPt$ mide la variabilidad de este desfase. El uso del valor absoluto en la definición de estas medidas hace énfasis en la importancia del error en la medición regional: errores de signos opuestos en regiones distintas siempre se suman.

Las medidas de error descritas se calculan con base en los resultados de los censos tal y como se obtienen. Es posible argumentar que, dado que se conoce la matriz de transferencias T , los errores se deberían calcular con base en ajustes de los resultados del censo que reflejen las migraciones esperadas. Esta observación es correcta, y lleva a una aclaración sobre la naturaleza de los flujos poblacionales registrados en T .

Existen distintos motivos para las migraciones de población. Algunos de ellos presentan patrones predecibles *ex ante*, como la migración laboral, en tanto que otros no. Los flujos que nos interesan en esta simulación son aquellos que no son predecibles. Es por esto que la matriz de transferencias T refleja los desplazamientos debidos a la violencia, pero no los laborales. Su función es dar

una medida de la magnitud de tales flujos no esperados en Colombia, sin pretender predecir el patrón que se observará en los siguientes años. No tiene entonces sentido ajustar las mediciones de los censos para reflejar los datos históricos en T , puesto que éstos no implican expectativas futuras.

3.1.3 Parámetros de la simulación

Las simulaciones de Monte Carlo (MC) tienen la ventaja de ser flexibles y permitir el modelaje de dinámicas complejas. Sin embargo, es difícil hacer análisis de estática comparativa con base en algoritmos MC. Es por este motivo que en este trabajo se presentan dos grupos de simulaciones con dos puntos de partida distintos.

El primer grupo de simulaciones tiene por objeto observar los efectos de los flujos poblacionales y su interacción con el crecimiento demográfico, aislando éstos efectos de aquellos debidos a la configuración inicial de la población. En estas simulaciones, el vector de población inicial es uniforme, con 10.000 individuos en cada región, y las matrices de transferencias son muy sencillas, diseñadas para generar dos tipos de migraciones: (i) ruido, es decir, migraciones que no generan flujos netos de población entre regiones, y (ii) flujos netos de población. Dentro de las matrices de ruido, se llevan a cabo simulaciones con ruido uniforme en todas las regiones y simulaciones en las que unas regiones tienen más ruido que otras.

El segundo grupo de simulaciones reproduce el caso colombiano. La matriz de transferencia corresponde a los datos históricos que se describen en la sección siguiente, y la población de cada región refleja la población en distintas regiones de Colombia.

Es importante recalcar que este segundo grupo de simulaciones es a escala. La población colombiana es 43'729.743 de personas según el censo de 1993, y el estimado de desplazados entre 1995 y 2005 es de 1'542.915. Sin embargo, nuestra matriz de transferencias se construye con base en encuestas a una muestra de los hogares desplazados incluidos en el Sistema RUT. Por lo tanto, es

necesario escalar bien sea la población total o la matriz de transferencias para que sean consistentes entre sí.

Debido a limitaciones computacionales en el cálculo de la distribución inversa de Poisson, la población total no puede utilizarse directamente para las simulaciones.

Por lo tanto se escala el vector de población total por un factor de $40102/1542915$.

Para apreciar el efecto de este escalamiento en los errores, se lleva a cabo adicionalmente una simulación con un factor de escalamiento diferente y se comparan los resultados.⁶

3.2. Descripción de los datos

Dos bases de datos son utilizadas para estimar los errores de conteo de ambos tipos de censo: el Sistema Único de Registro – SUR - de la Red de Solidaridad Social (RSS) y el Sistema RUT de la Conferencia Episcopal. El número de personas desplazadas y su distribución municipal se obtuvo del SUR. El propósito del SUR es identificar a las personas que se consideran legalmente desplazadas, caracterizar dicha población, cuantificar la demanda real por atención estatal y establecer un seguimiento a las ayudas prestadas por la RSS⁷. Para acceder al SUR, los hogares desplazados deben rendir una declaración frente a cualquier entidad del Ministerio Público, la cual debe ser valorada en un término de 15 días y, dadas ciertas condiciones, aceptar o negar su registro en el SUR. Los datos recolectados por el SUR no contienen información detallada acerca del proceso migratorio de los hogares, información esencial para construir la matriz de migración.

La matriz de migración se construyó, por consiguiente, con base en los datos del Sistema RUT. El Sistema de Información RUT se inició en 1997 y consta de un cuestionario aplicado a las familias desplazadas que se acercan a las Diócesis de la Iglesia Católica. El cuestionario contiene información acerca de las características sociodemográficas de los miembros del hogar, del acceso a

⁶ Los factores usados son $40.102/1'542.915$ y $(40.102/1'542.915)*0.5$.

⁷ Red de Solidaridad Social, www.red.gov.co, 16 de marzo de 2005.

mercados laborales y la asistencia educativa, de la tenencia de tierras y una información detallada del proceso migratorio. La información del proceso migratorio comprende el municipio de origen, el municipio receptor y los municipios de estadía provisional así como el tiempo de residencia en cada municipio. El Sistema RUT contiene información de 32.093 hogares desplazados correspondiente a 157.491 personas⁸.

Para construir la matriz de migración, se agruparon los municipios del país en 48 regiones. Previo a la selección de las regiones, se identificaron los patrones de migración de las familias desplazadas con el fin de conocer los destinos habituales para las familias de cada región. Las regiones se escogieron, por lo tanto, de tal manera que no se perdiera el proceso de migración en el momento de la agrupación y que se minimizara el porcentaje de ceros en cada una de las celdas de la matriz.

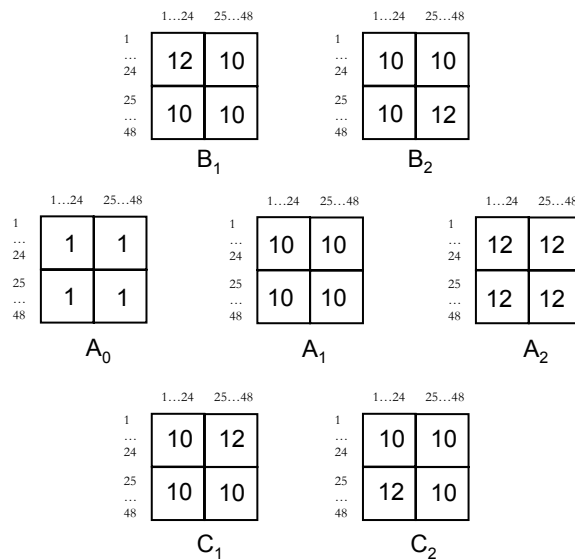
3.3. Discusión de resultados

3.3.1 Primer grupo de simulaciones

El primer grupo de simulaciones utiliza un vector de población inicial uniforme con 10.000 personas en cada una de las 48 regiones. Las simulaciones contemplan distintas tasas de crecimiento poblacional anual para cada una de las seis matrices de transferencia ilustradas en la figura 1:

⁸ Corte de datos: 30 de julio de 2003.

Figura 1. Matrices de transferencia utilizadas en la simulación



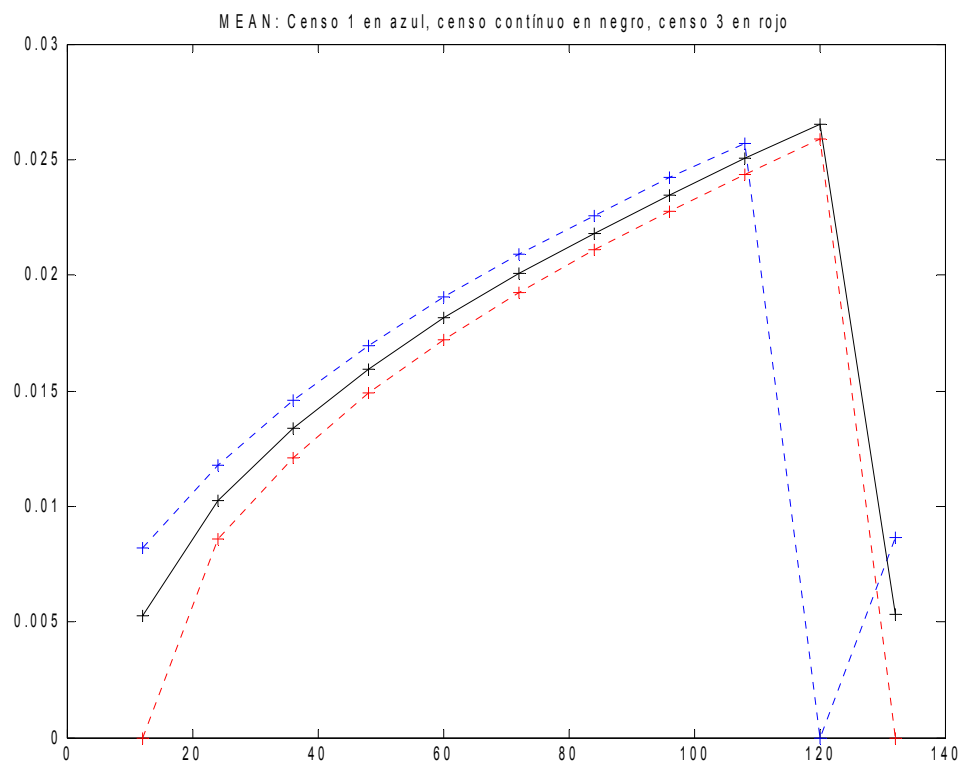
Las matrices A y B corresponden a flujos de población que no resultan en migraciones netas entre regiones ("ruido"). Para las matrices A, el ruido es homogéneo, en tanto que para las B es mayor en unas regiones que en otras. La matriz A_1 es la base de referencia para comparación en esta parte del trabajo.

Las matrices C implican flujos netos de población. C_1 corresponde aun flujo neto mensual promedio de 2 personas de cada una de las 24 primeras regiones a cada una de las 24 últimas. En C_2 el flujo neto es en la dirección opuesta.

Los flujos poblacionales presentan en principio dificultades particulares para el censo continuo. En el censo tradicional la medición inicial no tiene error en el conteo global de la población. En el censo continuo este no es el caso si hay flujos netos de población entre regiones durante la realización del censo. Por ejemplo, si la región que se censa en el primer periodo transfiere población a la región que se censa en el tercer periodo, algunas personas serán censadas dos veces y, por lo tanto, aparecerán dos veces en el conteo final. En este caso, la población total del país se sobreestima. Cuando la dirección de flujo va en sentido contrario la

población total se subestima. El problema es por supuesto mayor si los flujos netos son sistemáticamente diferentes de cero.

Figura 2. Evolución anual del error porcentual promedio de los censos bajo A_1



Fuente: Cálculo de los autores

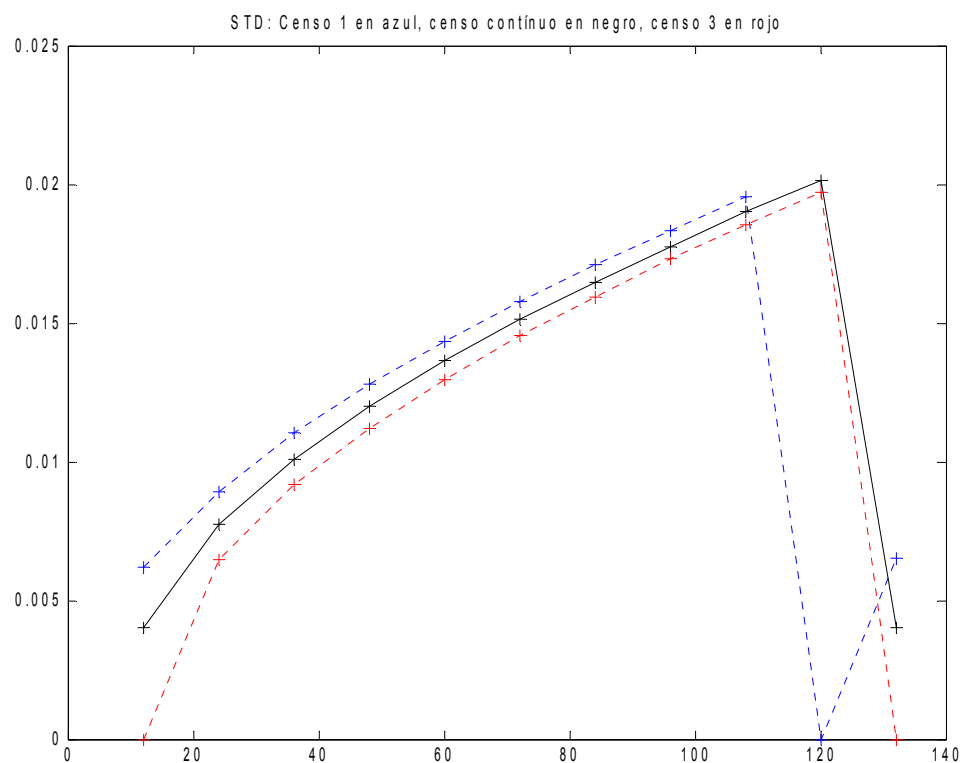
La figura 2 presenta la evolución del error promedio de los censos EPP_t bajo la matriz A_1 , que sirve como base de referencia para las comparaciones. La línea superior corresponde al primer censo tradicional, la del medio al censo continuo y la inferior al segundo censo tradicional. Como es de esperarse, los errores porcentuales de los tres censos aumentan a medida que pasa el tiempo y el momento de la medición de la población observada se aleja.

Los errores del censo continuo se encuentran entre aquellos del primer y el segundo censo tradicional. La diferencia en el error entre los dos censos tradicionales refleja la pérdida de exactitud de estos censos en el transcurso de un año. En el caso del censo continuo, el error en las regiones que se miden

temprano en el año se parece al del primer censo tradicional, y el de las últimas regiones se parece al del segundo censo.

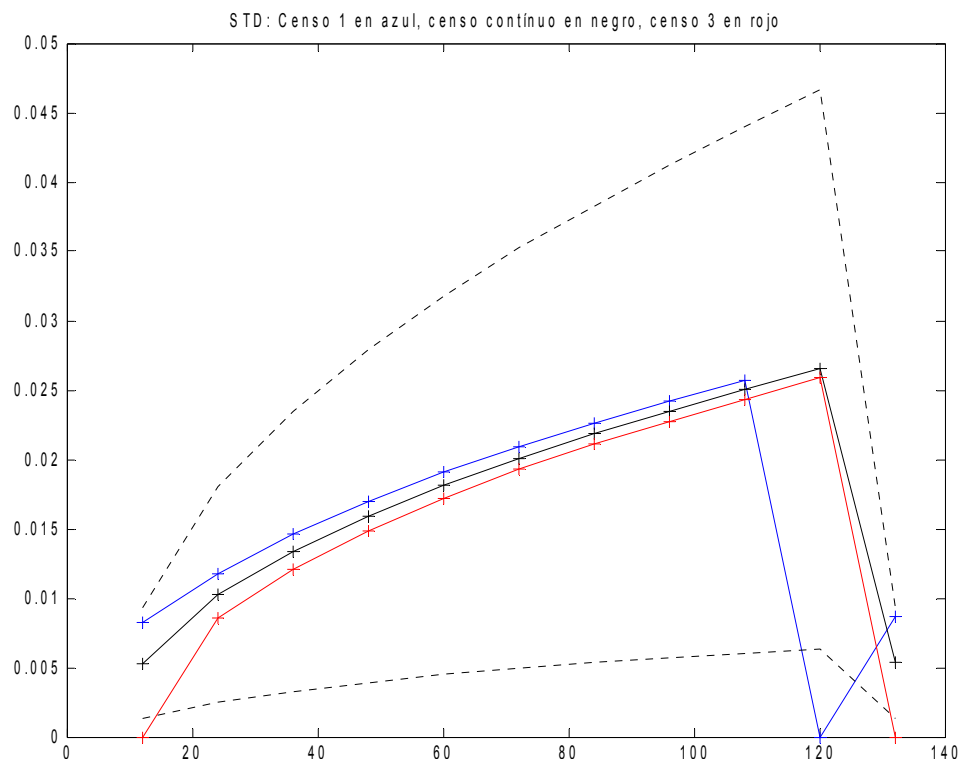
La figura 3 ilustra la evolución de las desviaciones estándar de los errores, DPP_t , bajo A_1 , y refleja la variabilidad del error. El orden de magnitud de los errores promedios y de sus desviaciones estándar es comparable entre censos, y los niveles de errores son indistinguibles estadísticamente. La figura 4 superpone los errores promedio de los tres censos (líneas continuas) y la desviación estándar promedio alrededor del censo continuo (líneas punteadas).

Figura 3. Evolución anual de la desviación estándar promedio en el error porcentual de los censos bajo A_1



Fuente: Cálculo de los autores

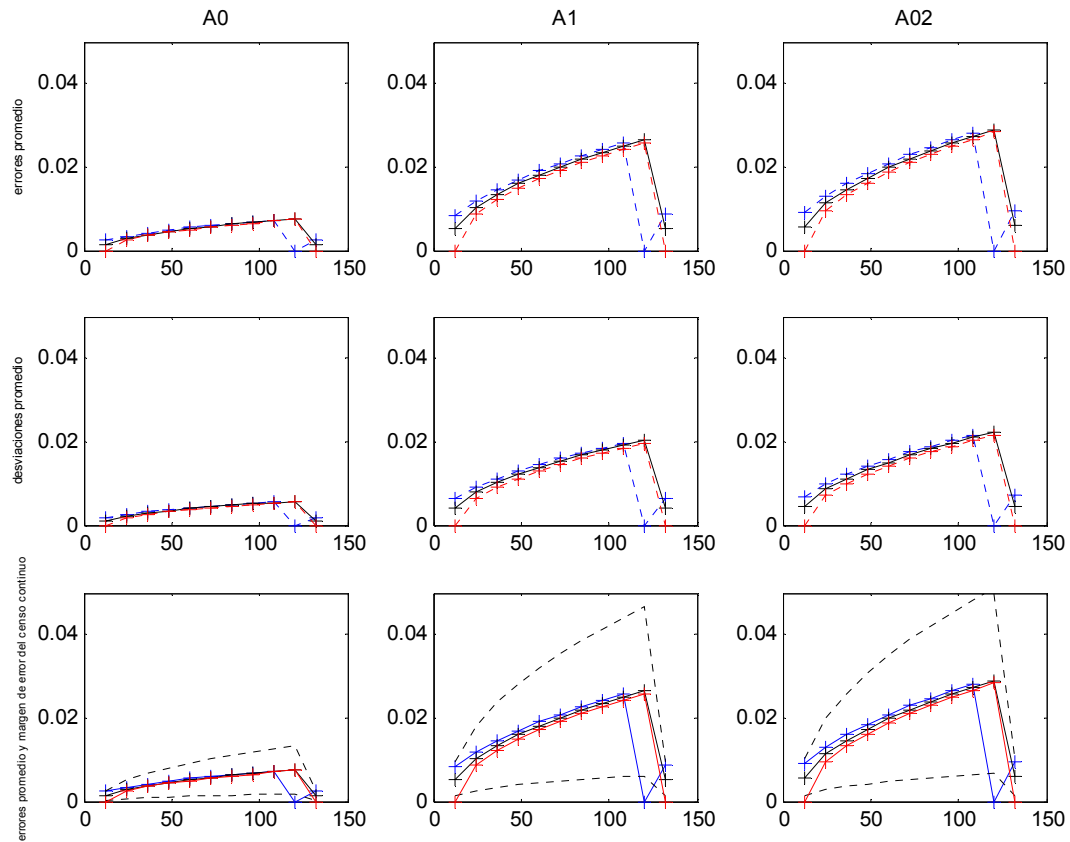
Figura 4. Evolución anual del error porcentual promedio de los censos y margen de error del censo continuo bajo A_1



Fuente: Cálculo de los autores

En este trabajo, las características de las curvas de error promedio que se consideran son su concavidad, el valor a los 120 meses y la dispersión entre los censos. En las figuras anteriores, estas características son similares para los tres tipos de censos. Veremos más adelante que ellas cambian según la matriz de transferencias y el crecimiento de la población. Sin embargo, dadas matriz y tasa de crecimiento, las curvas de los tres censos son siempre similares.

Figura 5. Evolución anual del nivel y la variabilidad del error porcentual promedio de los censos bajo las matrices A



Fuente: Cálculo de los autores

La figura 5 presenta el efecto del nivel de ruido homogéneo sobre EPP_t (fila superior) y DPP_t (segunda fila). La fila inferior presenta la desviación estándar alrededor de EPP_t para el censo continuo, y muestra claramente que los errores de los tres censos son estadísticamente indistinguibles. Los resultados corresponden a las matrices de transferencia A, y el ruido aumenta de izquierda a derecha. Tanto EPP_t como DPP_t aumentan con el ruido a tasas cada vez mayores, pero los resultados son siempre similares.

Las tablas 1a y 1b resumen los resultados de la comparación del efecto del ruido homogéneo sobre los censos. La tabla 1a presenta los niveles de error y variabilidad del error. El censo continuo presenta errores levemente mayores y

más variables que los censos tradicionales, pero la diferencia es insignificante en comparación con la variabilidad de los errores en todos los casos.

La tabla 1b resume la relación entre el error y el nivel de ruido. Entre A_0 y A_1 la elasticidad del error del censo continuo a los 120 meses con respecto al ruido es aproximadamente 0.349, mientras que entre A_1 y A_2 es 0.912. Para DPP_t , las elasticidades del mismo censo con respecto al ruido son 0.354 y 0.908 respectivamente.

Tabla 1a. Nivel del error a los 120 meses según nivel de ruido

T	Censo tradicional 1		Censo continuo		Censo tradicional 2	
	Promedio	Desv. Est.	Promedio	Desv. Est.	Promedio	Desv. Est.
A_0	0.0073	0.0056	0.0076	0.0057	0.0074	0.0056
A_1	0.0257	0.0195	0.0265	0.0202	0.0259	0.0197
A_2	0.0281	0.0213	0.0290	0.0220	0.0283	0.0215

Fuente: Cálculo de los autores

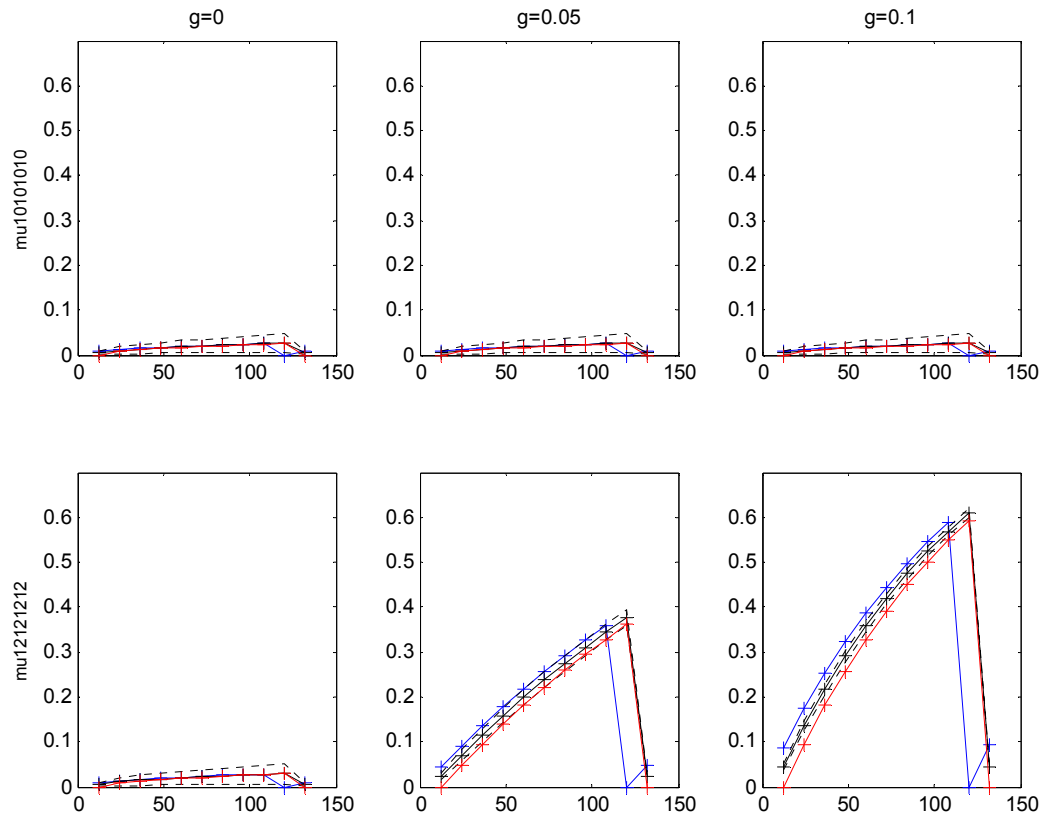
Tabla 1b. Elasticidad del error a los 120 meses con respecto al ruido

T	Censo tradicional 1		Censo continuo		Censo tradicional 2	
	Promedio	Desv. Est.	Promedio	Desv. Est.	Promedio	Desv. Est.
A_0 a A_1	0.352	0.348	0.349	0.354	0.350	0.352
A_1 a A_2	0.911	0.910	0.912	0.908	0.911	0.910

Fuente: Cálculo de los autores

La figura 6 reporta las curvas de error para distintos crecimientos poblacionales bajo las matrices A_1 (fila superior) y A_2 . Las líneas punteadas muestran la desviación estándar alrededor del censo continuo. A medida que aumenta la tasa de crecimiento, el nivel de error aumenta en todos los casos, pero el aumento es mucho mayor entre mayor es el ruido. El desempeño de los censos es similar.

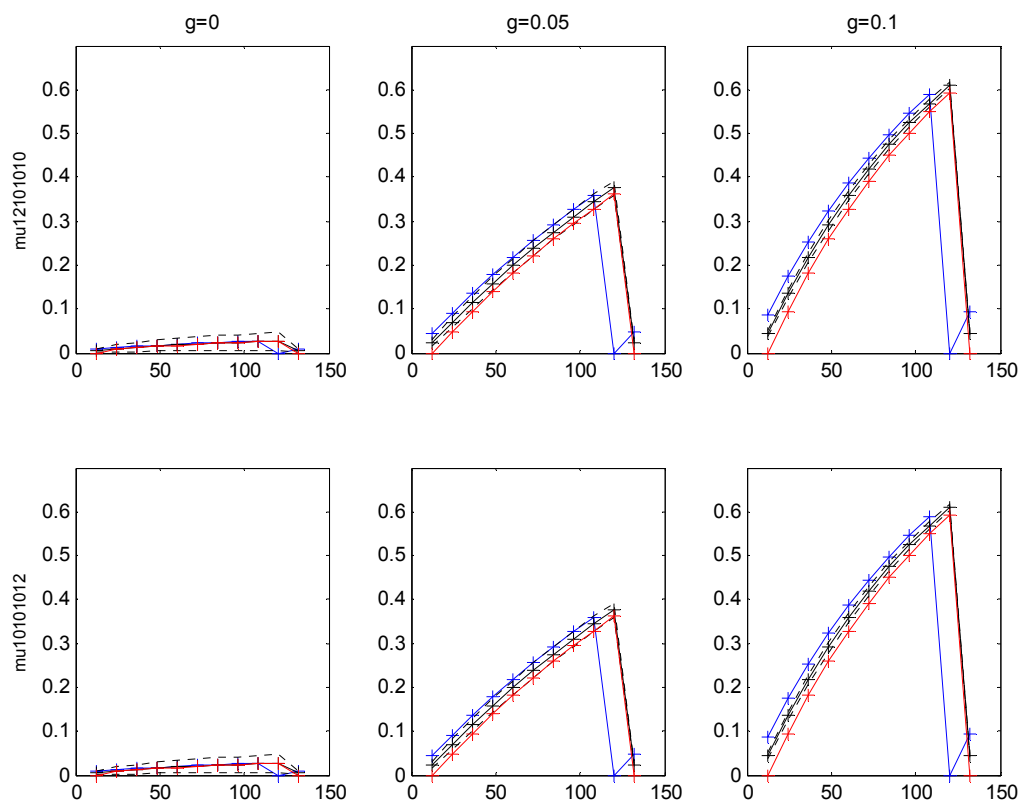
Figura 6. Evolución anual del error porcentual promedio de los censos bajo las matrices A para distintas tasas de crecimiento poblacional



Fuente: Cálculo de los autores

La figura 7 reporta las curvas de error para distintos crecimientos poblacionales bajo las matrices B_1 (fila superior) y B_2 , de forma que el ruido no es homogéneo. De nuevo, a medida que aumenta la tasa de crecimiento, el nivel de error aumenta en todos los casos, pero el desempeño de los censos es similar. Para cada tasa de crecimiento, el nivel de error de los censos está entre los registrados para las matrices A_1 y A_2 , y en el caso del censo continuo es más alto si las regiones con mayor ruido se censan temprano (matriz B_1) que si se censan tarde (B_2).

Figura7. Evolución anual del nivel y la variabilidad del error porcentual promedio de los censos bajo las matrices B para distintas tasas de crecimiento poblacional



Fuente: Cálculo de los autores

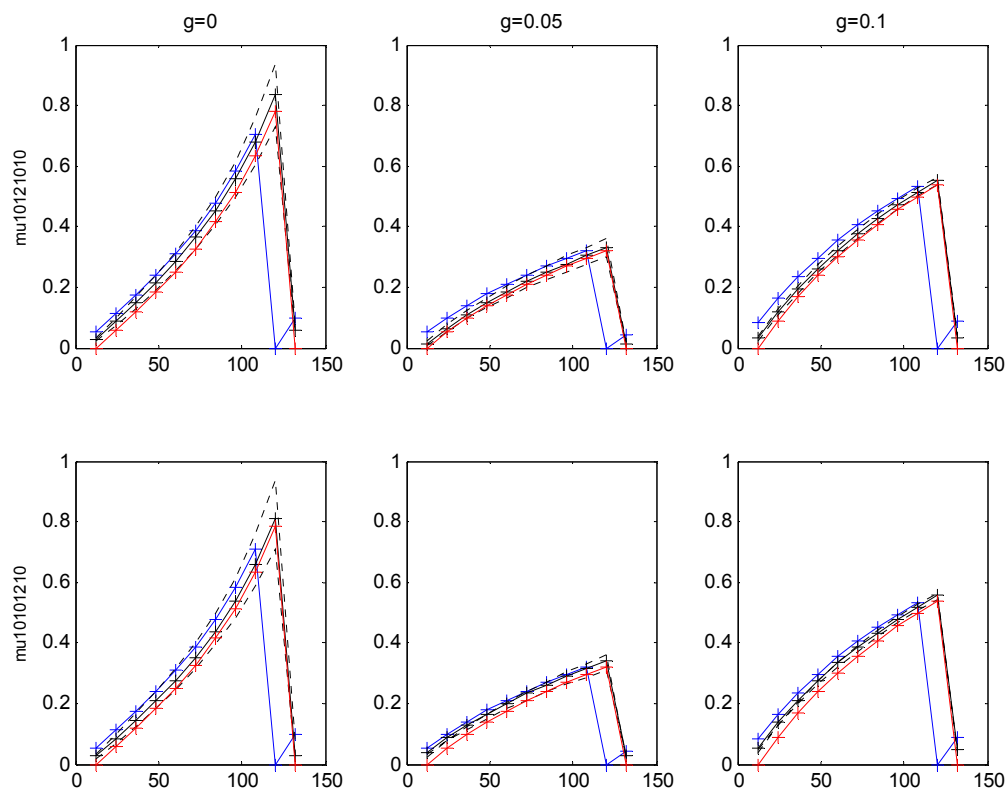
La figura 8 introduce matrices con flujos netos de población entre regiones. De nuevo, los censos se desempeñan similarmente. Sin embargo, para tasas de crecimiento poblacional bajas: (i) los errores son mayores, (ii) la concavidad de la curva de errores cambia, y (iii) el nivel máximo de error cae al aumentar la tasa.

Claramente, los flujos netos causan (i) y (ii) al hacer que la medición de una región dada se haga obsoleta más rápidamente. Es más, si bien las gráficas no lo registran,⁹ el censo continuo es asimétrico con respecto a las matrices C_1 y C_2 :

⁹ El motivo por el que las gráficas no registran la asimetría es que el error utilizado es el valor absoluto del error regional, de forma que no hay diferencia entre desviaciones hacia abajo y hacia arriba de la población real.

bajo C_1 , hay tendencia a contar personas doblemente y sobreestimar la población total, en tanto que bajo C_2 la tendencia es la opuesta. La característica (iii) se debe a que, al aumentar el crecimiento poblacional, los flujos netos de personas se hacen menores como porcentaje de la población total. La fuente de error dominante pasa a ser ese crecimiento, y como se ha visto la concavidad de la curva bajo crecimiento poblacional es hacia abajo.¹⁰

Figura 8. Evolución anual del nivel y la variabilidad del error porcentual promedio de los censos bajo las matrices C para distintas tasas de crecimiento poblacional



Fuente: Cálculo de los autores

¹⁰ En este trabajo no exploramos el efecto de crecimientos poblacionales heterogéneo sobre la concavidad de la curva de errores.

3.3.2 Segundo grupo de simulaciones

Para el segundo grupo de simulaciones, los datos son congruentes con las estadísticas colombianas. Las migraciones registradas entre 1995 y 2005 se describen en la tabla 2. La población descrita corresponde al total de la población colombiana según el censo de 1993. La muestra de desplazados, en cambio, corresponde a sólo 40.102 de un estimado de 1'542.915 entre 1995 y 2005. Como se ha descrito anteriormente, para las simulaciones se escaló la población por un factor de $40.102/1'542.915$ para hacer los datos compatibles.

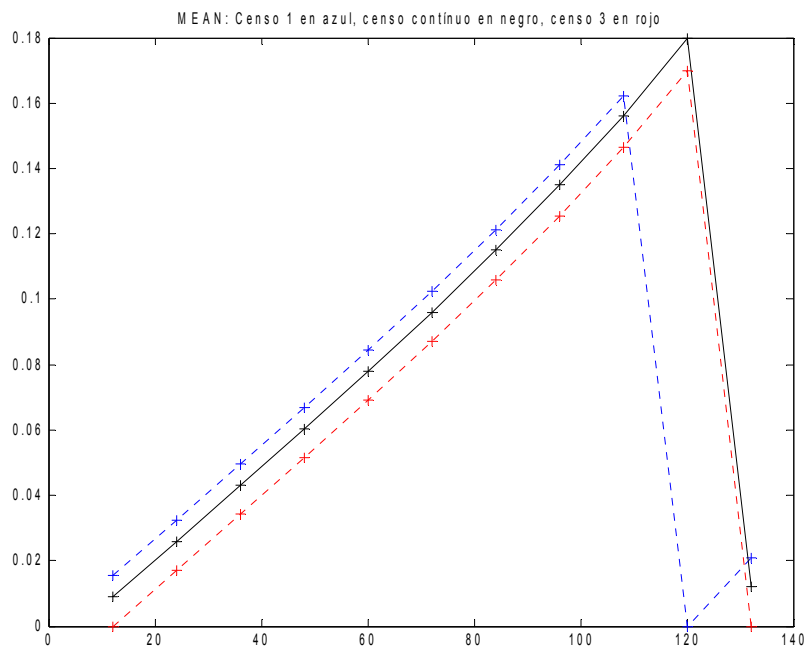
Las figuras 9 y 10 presentan la curva de errores y su variabilidad, respectivamente, para la simulación colombiana. El crecimiento poblacional en la simulación es 1.5% anual, consistente con estimaciones recientes. A diferencia de las simulaciones en el primer grupo, la variabilidad del error es un orden de magnitud menor que su nivel. Esto se aprecia en la figura 11, que superpone las curvas de errores y la variabilidad del censo continuo.

Tabla 2. Descripción de las migraciones en la matriz de transición para el caso colombiano

Región	Población (censo 1993)	Emigran	Emigrantes (%)	Inmigran	Inmigrantes (%)	Migración neta	Migración neta (%)
1	447'718	1'170	0.261%	3	0.001%	-1'167	-0.261%
2	3'470'682	1'608	0.046%	1'727	0.050%	119	0.003%
3	1'313'480	1'960	0.149%	1'229	0.094%	-731	-0.056%
4	1'655'615	5'067	0.306%	7'323	0.442%	2'256	0.136%
5	417'470	1'082	0.259%	265	0.063%	-817	-0.196%
6	187'844	1'937	1.031%	52	0.028%	-1'885	-1.003%
7	236'694	2'438	1.030%	3'629	1.533%	1'191	0.503%
8	739'808	817	0.110%	2'290	0.310%	1'473	0.199%
9	555'585	1'869	0.336%	473	0.085%	-1'396	-0.251%
10	471'556	426	0.090%	2'892	0.613%	2'466	0.523%
11	257'467	2'031	0.789%	242	0.094%	-1'789	-0.695%
12	813'680	186	0.023%	147	0.018%	-39	-0.005%
13	3'504'511	2'325	0.066%	2'811	0.080%	486	0.014%
14	135'662	902	0.665%	12	0.009%	-890	-0.656%
15	215'043	1'531	0.712%	25	0.012%	-1'506	-0.700%
16	666'101	1'873	0.281%	2'180	0.327%	307	0.046%
17	331'476	349	0.105%	2	0.001%	-347	-0.105%
18	222'464	2'034	0.914%	1'870	0.841%	-164	-0.074%
19	171'082	307	0.179%	0	0.000%	-307	-0.179%
20	123'560	1'642	1.329%	1'301	1.053%	-341	-0.276%
21	637'174	123	0.019%	161	0.025%	38	0.006%
22	715'105	1'132	0.158%	1'073	0.150%	-59	-0.008%
23	618'722	818	0.132%	980	0.158%	162	0.026%
24	205'946	283	0.137%	8	0.004%	-275	-0.134%
25	947'419	747	0.079%	236	0.025%	-511	-0.054%
26	366'673	126	0.034%	1	0.000%	-125	-0.034%
27	279'804	265	0.095%	50	0.018%	-215	-0.077%
28	1'394'504	557	0.040%	2'681	0.192%	2'124	0.152%
29	348'447	379	0.109%	169	0.049%	-210	-0.060%
30	1'666'143	313	0.019%	551	0.033%	238	0.014%
31	226'705	470	0.207%	1	0.000%	-469	-0.207%
32	907'085	261	0.029%	824	0.091%	563	0.062%
33	283'340	255	0.090%	12	0.004%	-243	-0.086%
34	1'121'957	174	0.016%	253	0.023%	79	0.007%
35	524'611	426	0.081%	819	0.156%	393	0.075%
36	428'815	158	0.037%	429	0.100%	271	0.063%
37	2'226'236	295	0.013%	848	0.038%	553	0.025%
38	1'299'256	537	0.041%	43	0.003%	-494	-0.038%
39	301'387	258	0.086%	133	0.044%	-125	-0.041%
40	976'964	247	0.025%	899	0.092%	652	0.067%
41	256'664	211	0.082%	70	0.027%	-141	-0.055%
42	88'899	138	0.155%	13	0.015%	-125	-0.141%
43	1'385'184	85	0.006%	189	0.014%	104	0.008%
44	500'029	120	0.024%	13	0.003%	-107	-0.021%
45	582'966	77	0.013%	219	0.038%	142	0.024%
46	26'573	68	0.256%	2	0.008%	-66	-0.248%
47	2'223'229	25	0.001%	952	0.043%	927	0.042%
48	7'252'408	0	0.000%	0	0.000%	0	0.000%
TOTAL	43'729'743	40'102	0.092%	40'102	0.092%	0	0

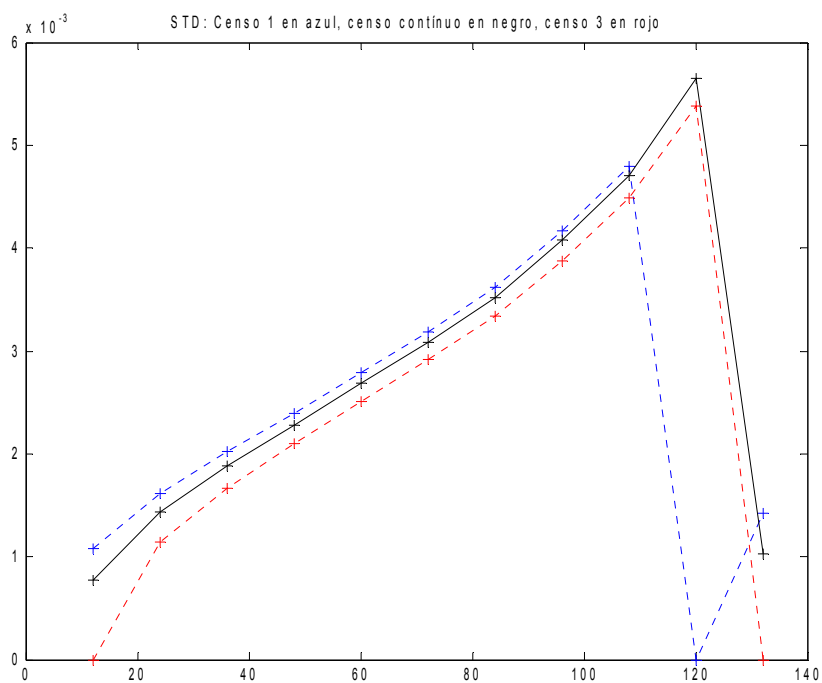
Fuente: Cálculo de los autores

Figura 9. Evolución anual del error porcentual promedio de los censos para Colombia



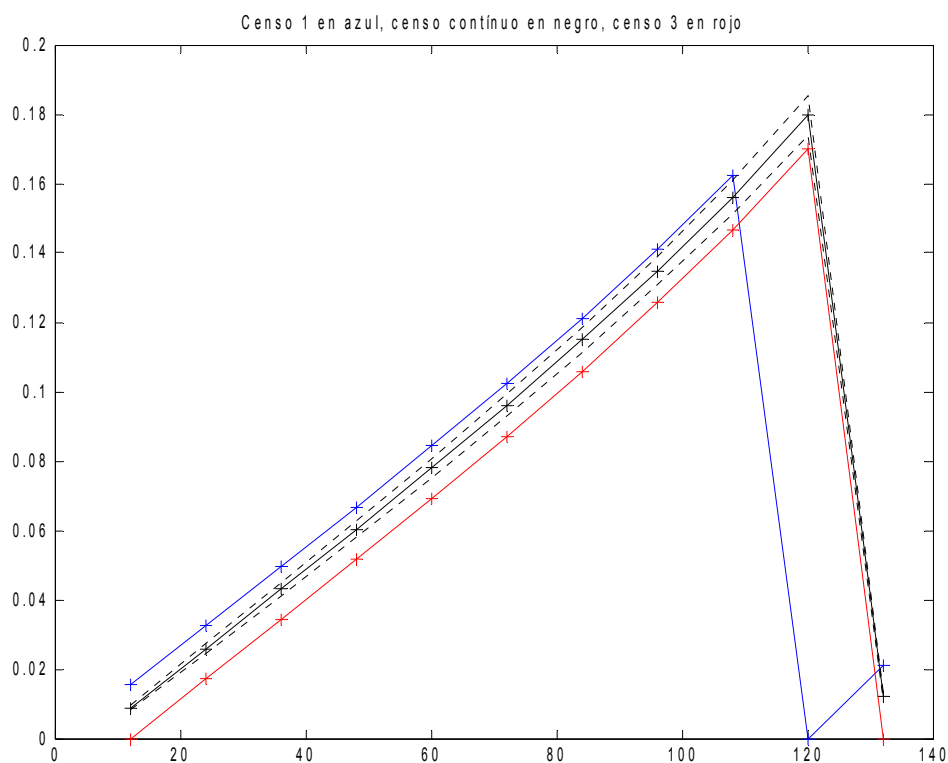
Fuente: Cálculo de los autores

Figura 10. Evolución anual de la desviación estándar promedio en el error porcentual de los censos para Colombia



Fuente: Cálculo de los autores

Figura 11. Evolución anual del error porcentual promedio de los censos y margen de error del censo continuo para Colombia



Fuente: Cálculo de los autores

La tabla 3 compara los niveles de errores tras 120 meses para los tres censos.

Tabla 3. Nivel del error a los 120 meses según nivel de ruido para Colombia

T	Censo tradicional 1		Censo continuo		Censo tradicional 2	
	Promedio	Desv. Est.	Promedio	Desv. Est.	Promedio	Desv. Est.
40.102/1'542.915	0.1624	0.0048	0.1797	0.0056	0.1698	0.0054

Fuente: Cálculo de los autores

Tras el análisis de las simulaciones en el primer grupo, es posible explicar las formas de las curvas en las figuras 9, 10 y 11. La concavidad de la curva refleja la presencia de flujos poblacionales netos entre regiones y de crecimiento poblacional, ninguno de los cuales domina en este sentido. Los flujos netos explican también el alto nivel de error. Finalmente, el alto error, junto con la

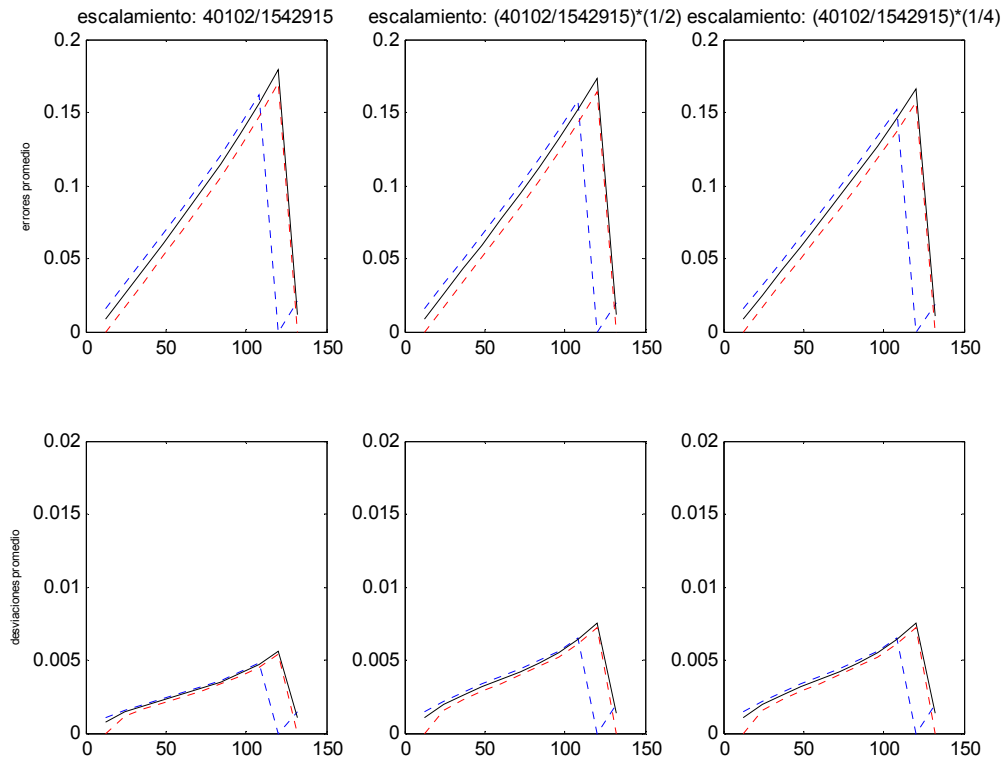
población total elevada (1'536.582 en la simulación) explican que la variabilidad del error sea pequeña en comparación con su nivel.¹¹

Para el caso colombiano, las curvas de error de los censos tradicionales no están dentro de una desviación estándar del censo continuo. Esto significa que, en cada momento del tiempo – digamos, al final del año cinco – el error del censo continuo es menor, en el sentido estadístico, que el del primer censo tradicional, y mayor que el del segundo. Sigue siendo cierto, sin embargo, que el deterioro de la exactitud del censo tradicional en el transcurso de un año es mayor que la diferencia en el error entre el censo tradicional y el continuo.

Finalmente, la figura 12 presenta el efecto del escalamiento de la población – y de las migraciones – bajo factores adicionales de $\frac{1}{2}$ y $\frac{1}{4}$ sobre las magnitudes relativas de los errores y las desviaciones estándar de los mismos. La primera fila corresponde a los errores porcentuales promedio, y la segunda a la desviación estándar de los mismos, y las columnas corresponden a distintos factores de escalamiento.

¹¹ Para las simulaciones del primer grupo los flujos mensuales eran entre 0.010% y 0.120% de la población en cada región. Para el caso colombiano, el valor promedio es del 0.059% pero varía según la región.

Fig.12: Distorsiones en los resultados debido al escalamiento de la población



Fuente: Cálculo de los autores

El error porcentual promedio no varía significativamente con el factor de escalamiento. La desviación estándar del error, por otro lado, aumenta al multiplicar la población y los flujos poblacionales por factores cada vez menores. Esto indica que los errores porcentuales regionales simulados son similares a los que se obtendrían con una simulación del escenario colombiano con escala 1:1, pero la desviación estándar simulada de los errores es mayor que la que se tendría para la simulación 1:1. En el caso extremo, este resultado sugiere que la estadística más informativa es el error porcentual promedio, y que en una simulación con una población de 44 millones de habitantes la variabilidad del error sería insignificante.

IV. Conclusiones e implicaciones de política

Las metodologías de censo continuo y tradicional difieren en varios aspectos que cubren desde el ámbito legal hasta su estrategia logística. El presente trabajo se enfoca en evaluar únicamente una de las diferencias entre los dos tipos de censo: la exactitud de la información obtenida y su degradación, es decir, la pérdida de exactitud en el conteo de la población colombiana a través del tiempo.

El análisis se basa en simulaciones de Monte Carlo de la evolución de la población de 48 regiones a lo largo de 132 periodos, entendidos como meses (once años). La población puede desplazarse entre las regiones y hay crecimiento demográfico. Durante el tiempo de la simulación se llevan a cabo seis conteos poblacionales, dos correspondientes a la metodología de censo continuo (durante doce meses) y cuatro correspondientes a censos tradicionales.

La variable analizada es el número de habitantes en cada región. Para comparar la exactitud de los censos, se calculan el promedio de los errores porcentuales en las regiones -- una medida de la magnitud del error -- y la desviación estándar promedio de esos errores -- la variabilidad del error.

Los distintos escenarios de simulación contemplados corresponden diversas tasas de crecimiento demográfico entre cero y diez por ciento, y diversos patrones de migración interregional. Finalmente se lleva a cabo una simulación calibrada al caso colombiano.

En todos los casos, los resultados muestran que ninguna de las metodologías de censos es significativamente superior a la otra en términos de exactitud. El error porcentual promedio del censo continuo es comparable en magnitud al del censo tradicional, y su evolución en el tiempo presenta el mismo patrón de degradación. Si se lleva a cabo un censo tradicional en el primer mes del censo continuo, y otro en el último mes, el error del censo continuo se mantiene entre los errores de ambos censos tradicionales. Esto equivale a decir que la diferencia entre la exactitud de un censo tradicional y un censo continuo es menor que la

degradación de la información del censo tradicional en el lapso de un año. Dado que la vida útil del censo en Colombia es de diez años, esta diferencia es menor.

La variabilidad del error porcentual promedio del censo continuo presenta también una evolución similar a la de los censos tradicionales. Es decir, las estimaciones de error tienen la misma confiabilidad bajo ambas metodologías de censo.

Un último resultado del estudio tiene que ver con la manera en que las migraciones interregionales inesperadas y el crecimiento poblacional afectan la exactitud de la información contenida en cualquiera de los censos. Se encuentra que los flujos poblacionales inesperados netos entre regiones generan una degradación cada vez más rápida de la información. En ausencia de flujos netos la degradación es cada vez más lenta.

Este estudio compara las metodologías de censos en un aspecto técnico específico, y encuentra que ninguna es superior a la otra. Sin embargo, la discusión de los méritos de una y otra metodología debe incluir consideraciones legales, logísticas y políticas que van más allá del alcance de este documento, y que posiblemente son de mayor relevancia en el contexto colombiano que los aspectos técnicos. En este sentido, la contribución del presente análisis es despejar el camino para que la discusión se concentre en esos aspectos.

Referencias

Ibáñez, A.M. y P. Querubin (2004). Acceso a tierras y desplazamiento forzado en Colombia. Documento CEDE 2004-23. Facultad de Economía. Universidad de los Andes. Bogotá, Colombia.

ONU - Organización de las Naciones Unidas (1997). *Principles and recommendations for Population and Housing Censuses*. Statistics Division. Department of Economic Affairs and Social Affairs.